Fault tree analysis - common misconceptions

Fault Tree Analysis – Common Misconceptions

Prof John Andrews, PhD.; Loughborough University, UK

Abstract

Fault Tree Analysis is a technique commonly used to assess the system failure likelihood or frequency in terms of failure and repair parameters of its components. It is a mathematical modelling method used to assess engineering systems in a very diverse range of industries. It has become apparent during my involvement in systems reliability quantification projects that there are several misconceptions involving the use of the method. These misconceptions range through all aspects of the fault tree method from the construction of the failure logic diagram through qualitative analysis and quantitative analysis. This paper deals with some of these misconceptions and the problems that can result.

Introduction

Many industries, particularly safety critical industries, use the fault tree analysis method (refs. 1, 3) to calculate the likelihood or frequency with which a system will fail in a particular mode. The method produces a failure logic diagram which appears as an inverted tree structure which defines the combinations of lower level component failure events which cause the top, system level failure. Its analysis is performed in two stages: the first is a qualitative analysis which identifies all of the minimal cut sets (necessary and sufficient combinations of component failures which cause the top event), the second then takes component failure data and quantifies the likelihood or frequency of system failure.

Misconceptions exist about this technique which have arisen for a variety of different reasons. Some are due to limitations which exist in the capabilities of commercially available software. Others are due to the fact that the methodology and mathematics behind the fault tree method are not fully understood. Aspects of the method may be true in many circumstances but are not the case in every circumstance. The misconceptions discussed in this paper cover the fault tree construction process, the modelling capabilities, analysis features and interpretation of the results.

Have you have heard any of the following types of statements or do they express your views on fault tree analysis?

- 'Fault Trees are only capable of modelling systems whose components have a constant failure rate'.
- 'The use of NOT logic in a fault tree is a mathematical complexity with no relevance to engineering systems analysis'.
- 'Fault Trees are a snap shot in time, they have no capability of modelling any sequential nature of failure events'.
- 'the weakest component is the one with the highest importance value'.

The views expressed in these statements will be discussed in this paper.

Fault Tree Construction

To provide a systematic way in which the construction of the fault tree may be approached, a method was proposed in the fault tree handbook (ref. 2) which was aimed at getting the logic structure correct. The approach required events in the fault tree development to be classified as state-of-system faults or state-of-component faults. A state-of-component fault is one which can be caused by a single component failure. If a single component failure cannot cause this then it is classified as a state-of-system fault. State-of-component faults are then developed using the fault tree structure illustrated in figure 1.

Originally published by the System Safety Society, in the Proceedings of the20th International System Safety Conference, held at Denver, August 2002

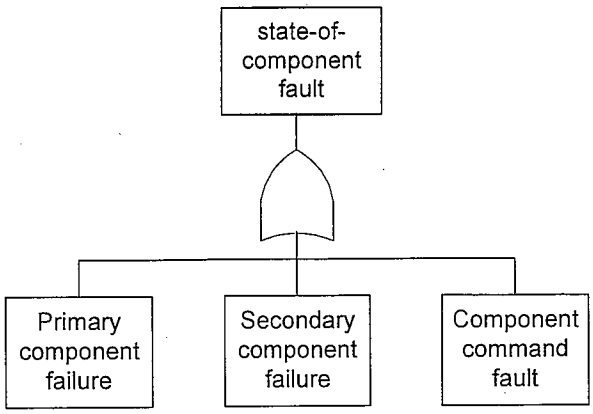

Figure 1 - State-of-Component Fault Tree Development

In figure 1 a **primary failure** is a component failure which occurs when the component is operating in its normal expected environment. A **secondary failure** is one where the component is operating outside its intended operating environment (usually due to other failures occurring which cause an increased stress level on the component). The **command fault** traces the fault back into other parts of the system which could cause a working component to exhibit the fault being developed (eg control system failures).

This approach has also been incorporated into texts which cover fault tree method in detail (ref. 1, 3). It is an effective way of generating a fault tree with the correct failure logic and is in itself non-controversial. The potential problem comes in the later analysis stage where all basic events in the fault tree structure are assumed to occur independently. The construction of the fault tree using this structure introduces dependencies between the basic events due to the effects a secondary failure has on minimal cut sets repair time. This can introduce large errors into the numerical procedures used to calculate the top event probability based on the assumption of independence.

As an example, consider the simple part of pressure tank system illustrated, along with the relevant section of the fault tree development, in figure 2. It is required to predict the unavailability of the pressure tank due to its rupture. This can be classified as a state-of-component fault since failure of the tank alone can produce this event. It is therefore developed as described above into its primary and secondary causes (in this example the tank does not have a command fault). The primary failure event is that the tank fails under normal expected conditions. The secondary failure event when the tank fails whilst operating outside its normal expected operating conditions is caused by an overpressure situation. The overpressure which ruptures the tank (assuming overpressure will always have this outcome) is due to the pump control system (represented as one basic event for simplicity) failing in such a way that the pump runs for too long AND the safety feature (the pressure relief valve) fails. The significant features of this situation are represented by the small, simplified, fault tree shown in figure 2. The minimal cut sets (minimal combinations of component failures which cause the system failure mode to occur) for this fault tree are:

1. TANK
2. PUMP. PRV

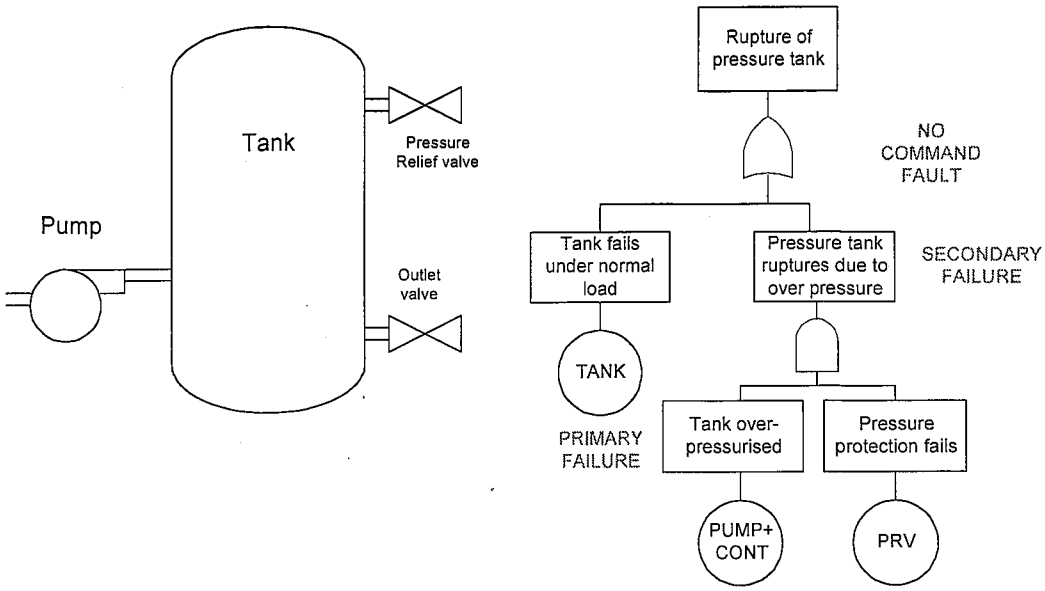

Figure 2 - Fault Tree for Simple Pressure Tank System

The implication of this is that by removing any of the events in the minimal cut sets the top event will not longer exist. However consider minimal cut set 2, if these two events occur together as well as these two components being in the failed state, the tank will also fail. This is not event in the minimal cut set. Considering the minimal cut set alone, if the pump or pressure relief valve are repaired it would under conditions of independence rectify the top event. However, since this failure is a secondary failure combination which will result in tank rupture the tank must also be repaired to rectify the system. Since the repair time for this component is likely to be considerably longer than the two elements of the minimal cut set failure of the analysis to account for this will result in a serious underestimate of the system unavailability.

The correct modelling of a section of the fault tree which features secondary failures would need to performed with a technique such as Markov methods (ref. 4) which can take into account the repair time dependence.

Fault Tree Analysis

Two common misconceptions which occur in the fault tree analysis stage of the modelling will be discussed separately. These are:

1. Fault Tree Analysis is a snap shot in time and cannot account for the sequence in which the basic events occur.
2. The use of Not logic in fault tree analysis is a mathematical complexity which has no practical engineering relevance.

Time Sequencing: Fault tree analysis, performed by traditional methods, makes the assumption that the order in which the basic component failures occur is irrelevant. However, situations frequently occur, particularly when modelling safety systems, that the order of events is vital to determine when event combinations will the cause the top event under consideration or not. The exact order of failure events is not important – which event is the last to fail is. Consider for example the simplified representation of a butane vaporising system shown in figure 3. Liquid butane is pumped from its storage tank to the vaporiser. The delivery pump has the capability to pump the butane at a pressure beyond the rating of the vaporiser coils. The protection against high pressure butane entering the vaporiser is provided by three safety features: two trip loops and a vent valve. The trip loops close on high pressure detection, the vent valve opens at high pressure and allows the butane to flow back to the tank. A fault tree which is developed for the top event 'butane vaporiser ruptures due to high pressure' would produce the failure mode (minimal cut set):

Pump. TR1 .TR2 .VV

Where 'Pump' indicates a pump surge, 'TR' denotes failure of each of the trip loops and 'VV' is the vent valve failure.

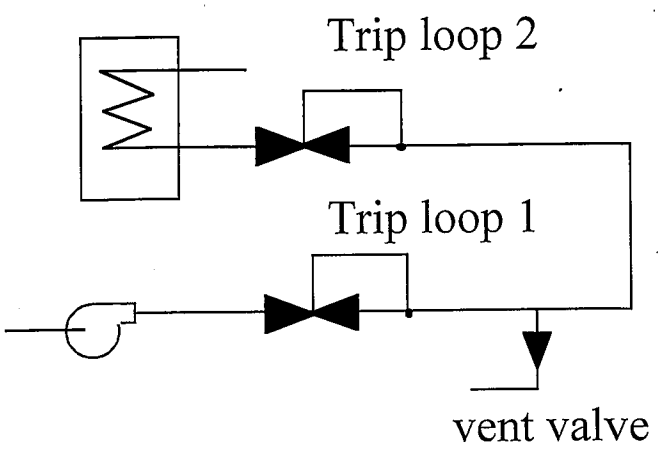

Figure 3 - Butane Vaporiser System

It is assumed that only one event can occur in an infinitesimally small time interval so for a minimal cut set to occur during a time interval [t,t+dt] it requires all events except one to have occurred by time t and the remaining event, the ith component failure, happens at time t. If order doesn't matter then each of the events in the cut set can in turn be considered the last to fail. In which case the frequency of failure for the cut sets above, w_{sys}, is given by:

$$w_{sys} = \sum_{i=1}^{n} (w_i(t) \prod_{\substack{j=1 \\ j \neq i}}^{n} q_j(t)) \qquad (1)$$

$$= w_{pump} q_{vv} q_{tr1} q_{tr2} + w_{vv} q_{pump} q_{tr1} q_{tr2} + w_{tr1} q_{vv} q_{pump} q_{tr2} + w_{tr2} q_{vv} q_{tr1} q_{pump}$$

where w is the unconditional failure intensity and q is the unavailability of each of the n components.

On a more detailed examination it can be seen that the only one of the four terms in equation 1 which actually causes the top event developed is the first. The others all have the hazardous event (pump surge) occurring before all the safety devices have failed and so a safe shut-down should occur. To get this calculation correct more attention must be focussed on the order in which the events fail. This can be accomplished by considering which events are initiators and which are enablers (ref. 5). These are defines as:

Initiating Events: Perturb system variables and place a demand on control/protection systems to respond.
Enabling Events: Are inactive control/protection systems which permit an initiating event to cause the top event.

These events are represented in figure 4 where the system failure mode occurs when the system is in a critical state (safety systems failed – enablers) and then the potential hazard occurs (initiator).

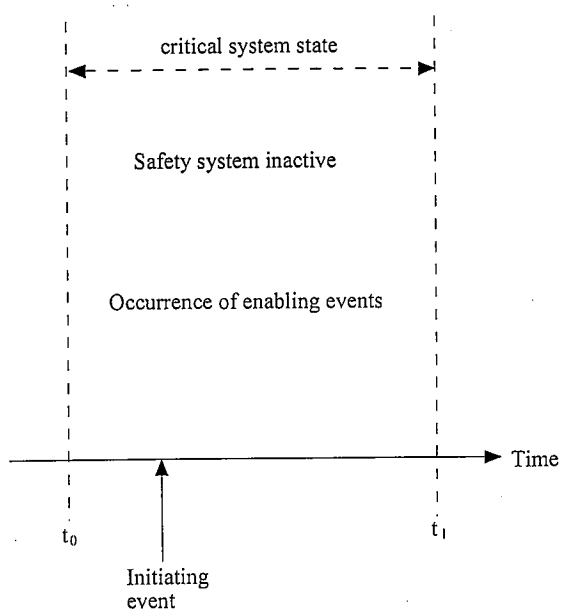

Figure 4 - Initiating and Enabling Events

For the vaporiser system the event which places a demand on safety systems to respond, initiator, is pump, safety system failure, enablers, are vv, tr1 and tr2. Considering the order of events, equation 1 becomes:

$$w_{sys} = \sum_{i=initiators} \left(w_i(t) \prod_{\substack{j=1 \\ j \neq i}}^{n} q_j(t)\right) \qquad (2)$$

$$= w_{pump} q_{vv} q_{tr1} q_{tr2}$$

As can be seen failure to account for initiators and enablers can result in a significant over-estimation of the failure frequency. A pessimistic assessment of the system but may result in resources being allocated incorrectly to this problem.

Not Logic: Most fault trees are constructed using AND and OR gates, if the third fundamental logic operator the NOT gate is used then the fault tree can become non-coherent. Non-coherent fault tree structures are generally discouraged by fault tree practitioners for both philosophical and practical reasons (ref. 6). Philosophically because the minimal combinations of events which cause the top event, termed prime implicants, contain component working states. So improving the state of the components in the system can make the system state deteriorate. The practical objections come because:

1. the prime implicants tend to be of larger order of more in number than the minimal cut sets of the corresponding coherent fault tree, increasing the size of the problem to be solved.
2. producing the complete list of prime implicants is a more complex analytical problem than producing minimal cut sets.
3. Approximations required in the conventional quantification process may not be valid.

Because of these reasons it is viewed by some as a mathematical complexity which last no practical significance to engineering problems. However most systems are non-coherent. Is there an argument that would justify that the inclusion of not logic should at least be considered? Consider the simplified system example illustrated in figure 5.

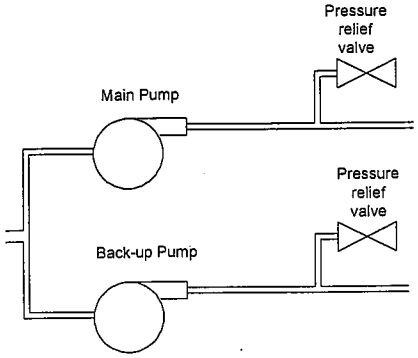

Figure 5 - Flow Control System

The flow of a hydrocarbon fluid is usually provided by the main pump which is protected by a pressure relief valve. In the event of a failure or maintenance taking place on this pump the active stream is switched to the backup. Consider the flowing sequence of events.

1. Flow is provided by the main pump and the backup is made non-functional for maintenance – PUMP-BU-F
2. The pressure relief valve on the backup line is removed for maintenance and blanked off but not made leak tight – PRV-BU-LEAK.

PUMP-BU-F and PRV-BU-LEAK together are two failure events which do not cause a problem.

3. the main pump fails, PUMP-M-F and the backup pump is restored, PUMP-BU-W to the functioning state to provide the flow requirement by staff not aware that the pressure relief valve has been removed.

The event combination: PUMP-M-F and PUMP-BU-W and PRV-BU-LEAK includes a working component state and releases the hydrocarbon to the environment where it may encounter an ignition source. This is a worse state than if the working component the backup pump was failed!

The above situation was that which occurred on the Piper Alpha offshore platform and resulted in the loss of 167 lives. This disaster happened because of a non-coherent combination of events. It certainly has relevance to many engineering systems and its exclusion in a systems analysis needs to be justified by stronger arguments than that it results in an increase in the mathematical complexity. Passive components such as pipes or wires frequently contribute to catastrophic system failures by transporting a disturbance in a system parameter into another part of the system where the hazard materialises i.e. the hazard is a result of a non-coherent combination.

Component Models

The probability of component failure events is determined by mathematical models which take account of the distributions of the time to failure and time to repair. It is a widely held view that the fault tree method is only capable of using models which feature a constant failure rate i.e. that the component is in its useful life period when no mechanisms such as wear, fatigue, corrosion or oxidation are influential. In such circumstances the time to failure distribution is the negative exponential form. The analyst therefore has to determine the distribution type for the repair times, decide if the failure is revealed or unrevealed and apply the correct model. If the failure is revealed, has constant failure rate λ and constant repair rate ν, its unavailability, q, is given by:

$$q(t) = \frac{\lambda}{\lambda + \nu}\left(1 - e^{-(\lambda+\nu)t}\right)$$

It is a falsity that the method can only deal with components with constant failure rates. If the failure or repair times are other than those from distributions with constant rates then analytical expressions such as that above are either difficult or impossible to obtain. Most of the commonly used commercial software packages available will only allow this assumption as clients do not request any other options. However any distributions can be used but the equations need to be solved numerically. If the times to failure have density function, f(t) and the repair times have density function, g(t) then the unconditional failure and repair intensities w(t) and v(t) are obtained from the simultaneous equations:

$$w(t) = f(t) + \int_o^t f(t-u)v(u)du$$

$$v(t) = \int_0^t g(t-u)w(u)du$$

the unavailability, q(t) is then given by:

$$q(t) = \int_0^t [w(u) - v(u)]du$$

It is not a limitation of the fault tree analysis technique that constant failure rates are assumed. It is commonly a limitation of the software implementation.

Results Interpretation

The fault tree analysis method will make predictions of the system failure parameters such as failure probability and failure frequency. If these predictions indicate that the system performance is inadequate then improvements need to be made. Importance measures indicate the contribution that each component makes, in some sense, to the system failure and can be used to determine the weakness in the analysed design. The component with the highest importance measure has the highest contribution. But the contribution to what depends on which of several possible importance measures are being calculated. Each measure has a different interpretation and this must be taken into account when interpreting the results – it is not correct to just say that the components with the highest importance ranking are the 'weakest'.

For example consider a 2-out-of-3 voting system with components A, B and C. The system failure minimal cut sets are therefore: {A,B},{A,C} and {B,C}. The failure and repair data, in hours, for these components are given in table 1.

Table 1 - Component Data

component	Failure rate λ	Mean time to repair (=1/repair rate)	probability
A	$1.x10^{-3}$	10	$1.x10^{-2}$
B	$1.x10^{-4}$	200	$2.x10^{-2}$
C	$5.x10^{-6}$	50	$2.5x10^{-4}$

Assuming that each component can act as either an initiator or an enabler the unavailability of the system is 2.0756×10^{-6} and the failure frequency is 2.137×10^{-5} per hour. The importance values for A, B and C for different importance measures are given in table 2. A definition of each measure can be found in references 1 and 3.

Table 2 - Importance Values

component	structural	Birnbaum	criticality	Fussell-Vesley	Initiator	Enabler
A	0.5*	0.0202	0.974	0.976	0.945*	0.050
B	0.5*	0.0102	0.983*	0.988*	0.048	0.940*
C	0.5*	0.0296*	0.036	0.036	0.007	0.023

* - the highest ranked event.

As can be seen depending which importance measure is selected each of the components A, B or C can be the most significant contributor. The way in which each of these measures is obtained is required to interpret the results.

The **structural measure** is a deterministic measure which takes no account of the probability of component failure. It only takes account of the system structure. In a 2-out-of-3 system the components all play an equal role in the structure.

The **Birnbaum measure** of importance is the probability that the system is critical for that component. That is a state such that the failure of the ranked component will fail the system. Component C has the highest contribution in this category. It is critical for component C (when C occurs is causes the system to fail) when either of components A or B have failed. Since A and B are the two highest probability components to fail, C will have the highest chance of the system being critical to it. To reduce the contribution of components indicated by this measure we could change the structure of the system or reduce the chances of failure for other components which make the system critical for the component. This measure is not a function of the components own failure probability.

The **criticality measure** is the chance that the system is critical for each component and that the component then fails (divided by the system failure probability). This is a function of the components own unavailability and so anything which reduces a components failure likelihood, such as getting a better quality component or reducing the repair time, will reduce the contribution signified by this measure.

The **Fussell-Vesley measure** of importance is the probability of the union of minimal cut sets containing the basic event divided by the system failure probability. This is numerically very similar to the criticality measure and component contributions can be reduced in the same way.

The final two measures of importance relate to the system failure frequency rather than the system failure probability. They rank separately the initiators and enablers.

The **initiator measure** of component importance ranks the contribution of each initiator. It is calculated as the probability that the system is critical to the failure of the initiator, multiplied by the failure frequency of the initiator ranked as a contribution to the expected number of system failures. Since each system failure can have only one initiator, adding these measures for each initiator will give unity. To reduce the ranking of any initiator in this list it is the failure frequency which plays a dominant role. Alternatively the criticality (as indicated by Birnbaum's measure) can also be reduced. As can be seen here the top ranked initiator is A which occurs most frequently.

The **enabler measure** ranks the contribution of components which allow other events to cause the top event. It is the probability of all the initiators (which appear in a minimal cut set along with the ranked component) cause the system failure when the enabler being ranked has also failed (again expressed as a contribution to the expected number of system failures). A component contribution by this measure can be reduced by reducing the unavailability.

Conclusions

1. If the most benefit is to be gained from the application of the fault tree analysis to engineering systems then it is necessary to have a good understanding of the method and not use it as a 'black box' piece of software.
2. It has been demonstrated that, failure to understand the method fully can result in an incorrect failure diagram, misinterpretation of the results, or an unnecessarily restricted analysis.

References

1. Andrews J.D,and Moss T.R., Risk and Reliability Assessment, Longman, 1993.
2. Vesely, W.E., Goldberg, F.F., Roberts, N.H. and Hassl, D.F., Fault Tree Handbook, Systems and reliability Research office of Nuclear Regulatory Research, US Nuclear Regulatory Commission, Washington DC, 20555, NUREG-0492, 1981.
3. Henley E.J and Kumamoto H., Reliability Engineering and Risk Assessement, Prentice Hall, 1981.
4. Hoyland, A and Rausand M, System Reliability Theory: Models and Statistical Methods, Wiley, 1994.
5. Lambert HE., Measures of Importance of Event and Min Cut Sets in Fault Trees, Reliability and Fault Tree Analysis, SIAM, Philadelphia, 1975.

6. Andrews J.D., The use of Not Logic in Fault Tree Analysis, Quality and Reliability Engineering International, Vol 17, 2001, pp143-150.

Biography

John D Andrews; Department of Mathematical Sciences; Loughborough University; Loughborough, Leicestershire, LE11 3TU, U.K. e-mail - J.D.Andrews@lboro.ac.uk.

Prof. J.D. Andrews is a Professor in the Department of Mathematical Sciences at Loughborough University. He joined this department in 1989 having previously gained nine years industrial research experience with British Gas and two years lecturing experience in the Mechanical Engineering Department at the University of Central England. His current research interests concern the assessment of the safety and risk of potentially hazardous industrial activities. This research has been heavily supported by industrial funding. Over recent years grants have been secured from the MOD, Rolls Royce Aero Engines, Mobil North Sea, and Bechtel. Professor Andrews has numerous journal/conference publications along with a jointly authored book "Risk and Reliability Assessment" which is now in its second edition.